A Characterization of Spherical Distributions

by

Morris L. Eaton[1]
University of Minnesota

Technical Report No. 430

[1] A portion of this work was completed while the author was a Visiting Professor at the University of California, Berkeley. This reasearch was supported in part by the National Science Foundation Grant MCS 81 00762.

Abstract

It is shown that when the random vector X in R^n has a mean and when the conditional expectation $E(u'X|v'X) = 0$ for all vectors $u, v \in R^n$ which satisfy $u'v = 0$, then the distribution of X is orthogonally invariant. A version of this characterization is also established when X does not have a mean vector.

Key Words: Orthogonally invariant distributions, spherical distributions, elliptical distribution, characterization, conditional expectation, error distributions, linear models.

AMS 1980 subject classifications: Primary 62E10, 62H05; Secondary 62J02.

1. Introduction and Summary

Recently, Toyooka (1982), Kariya and Toyooka (1983) and Eaton (1983) have discussed versions of the Gauss-Markov Theorem when the error covariance matrix is estimated. To motivate the results in this paper, we begin with a brief discussion of the type of situation treated in the three papers above. Consider a random vector Y in R^n with a mean vector μ and a covariance matrix $\Sigma \equiv \mathrm{Cov}(Y)$. The mean vector μ is assumed to lie in a known linear subspace $M \subseteq R^n$ and Σ is assumed to lie in a known set γ of positive define matrices (the assumption of positive definiteness can be dispensed with, but it is simpler for the discussion here to assume positive definiteness). Let P_Σ denote the unique projection with range M and null space $\Sigma(M^\perp)$ where $M^\perp$ denotes the orthogonal complement of M in R^n (with the usual inner product). Let A be all nxn matrices A which satisfy $A\mu = \mu$ for all $\mu \varepsilon M$. Hence AY is an unbiased estimator for μ for each $A \varepsilon A$. Note that $P_\Sigma \varepsilon A$. For each $A \varepsilon A$, $\mathrm{Cov}(AY) = A\Sigma A'$ when $\mathrm{Cov}(Y) = \Sigma$. One way to phrase the Gauss-Markov Theorem is that

$$P_\Sigma \Sigma P_\Sigma' \le A\Sigma A' \tag{1.1}$$

for each $A \varepsilon A$ where $\le$ refers to the Loewner ordering on symmetric matrices. In other words, the covariance matrix of $P_\Sigma Y \equiv \hat{\mu}_\Sigma$ is no larger than the covariance matrix of any other linear unbiased estimator of μ.

In many situations, Σ is not known so P_Σ, and hence $\hat{\mu}_\Sigma$, cannot be calculated without first estimating Σ. Estimators of Σ, say $\tilde{\Sigma} \varepsilon \gamma$, which satisfy

$$\begin{aligned} \tilde{\Sigma}(y) &= \tilde{\Sigma}(-y) \quad , \quad y \varepsilon R^n \\ \tilde{\Sigma}(y+x) &= \tilde{\Sigma}(y) \quad , \quad y \varepsilon R^n,\ x \varepsilon M \end{aligned} \tag{1.2}$$

are called <u>residual type estimators</u> (see Eaton (1983), Definition 6.1).

Most estimators of Σ that have been proposed in the literature are of this type. For such an estimator, let $\tilde{P} = P_{\tilde{\Sigma}}$ and let $\tilde{\mu} = \tilde{P}Y$. Thus, $\tilde{\mu}$ is just the Gauss-Markov estimator of μ using an estimated covariance matrix. To discuss properties of $\tilde{\mu}$, write the linear model for Y as

$$Y = \mu+\epsilon \quad , \quad \mu\varepsilon M \tag{1.3}$$

where the error vector ϵ has mean 0 and $\Sigma = \mathrm{Cov}(\epsilon)\varepsilon\gamma$. When ϵ and $-\epsilon$ have the same distribution, it is easy to show $\tilde{\mu}$ is an unbiased estimator of μ (for example, see Eaton (1983)). One question of interest is when does the Gauss-Markov Theorem hold--that is, when does the inequality

$$\mathrm{Cov}(\tilde{\mu}) \geq \mathrm{Cov}(\hat{\mu}_\Sigma) \tag{1.4}$$

hold for each $\Sigma\varepsilon\gamma$? One attempt to establish (1.4) is to let $Q_\Sigma = I-P_\Sigma$ and write

$$\tilde{\mu} = \tilde{P}Y = \tilde{P}P_\Sigma Y + \tilde{P}Q_\Sigma Y = P_\Sigma Y + \tilde{P}Q_\Sigma Y = \hat{\mu}_\Sigma + \tilde{P}Q_\Sigma Y. \tag{1.5}$$

Now, fix $\Sigma\varepsilon\gamma$. $\tilde{P}$ is a function of $\tilde{\Sigma}$ which in turn is a function of $Q_\Sigma Y$ since $\tilde{\Sigma}$ is a residual type estimator. Since $P_\Sigma Y$ and $Q_\Sigma Y$ are uncorrelated, there is some hope that $\hat{\mu}_\Sigma$ and $\tilde{P}Q_\Sigma Y$ might be uncorrelated. When this is the case, we have

$$\mathrm{Cov}(\tilde{\mu}) = \mathrm{Cov}(\hat{\mu}_\Sigma) + \mathrm{Cov}(\tilde{P}Q_\Sigma Y) \tag{1.6}$$

and hence (1.4). A useful sufficient condition that $\hat{\mu}_\Sigma$ and $\tilde{P}Q_\Sigma Y$ be uncorrelated is that the conditional expectation equation

$$\mathcal{E}(P_\Sigma\epsilon | Q_\Sigma\epsilon) = 0 \tag{1.7}$$

holds (see Eaton (1983), Section 7).

The main concern in this paper is equation (1.7) and its implications concerning $\mathcal{L}(\epsilon)$ - the distributional law of ϵ. First observe that it is sufficient to take $\Sigma = I$ in (1.7). To see this, write

$$Q_\Sigma\epsilon = \Sigma^{\frac{1}{2}}\Sigma^{-\frac{1}{2}}Q_\Sigma\Sigma^{\frac{1}{2}}\Sigma^{-\frac{1}{2}}\epsilon \tag{1.8}$$

and let $X = \Sigma^{-\frac{1}{2}}\varepsilon$ and let $Q = \Sigma^{-\frac{1}{2}}Q_{\Sigma}\Sigma^{\frac{1}{2}}$. Then X has mean 0, covariance I, and Q is the orthogonal projection onto $M^{\perp}$. Since $\Sigma^{\frac{1}{2}}$ is non-singular, the σ-algebra generated by QX and that generated by $\Sigma^{\frac{1}{2}}QX$ are the same. Hence conditioning on QX is equivalent to conditioning on $\Sigma^{\frac{1}{2}}QX$. Noting that $P = \Sigma^{-\frac{1}{2}}P_{\Sigma}\Sigma^{\frac{1}{2}}$ is the orthogonal projection onto M, we see that (1.7) is equivalent to

$$E(PX|QX) = 0 \qquad (1.9).$$

Recall that a random vector $Z \varepsilon R^n$ has a <u>spherical</u> <u>distribution</u> if $L(Z) = L(gZ)$ for all nxn orthogonal matrices g. It follows easily from results in Cambanis, Hwang and Simons (1981) that if X has a spherical distribution and if EX exists, then (1.9) holds for any subspace M where P is the orthogonal projection onto M and Q = I-P. In this paper we establish the converse--namely, if X has a mean and (1.9) holds, then X has a spherical distribution. In fact, we prove a bit more. To motivate the statement of our main result, take the subspace M to have dimension n-1 in R^n. Then $M^{\perp}$ has dimension 1 so $M^{\perp}$ = span {v} for some fixed vector $v \neq 0$. Since $Q = (vv')/v'v$, $QX = (v'X)(v/v'v)$ so conditioning on QX is equivalent to conditioning on v'X. Also, for any $u \varepsilon M$, Pu=u so (1.9) implies that

$$E(u'X|v'X) = 0 \qquad (1.10)$$

for all vectors u which are perpendicular to v. Here is our main result.

<u>Theorem 1</u>: Suppose $X \varepsilon R^n$ has a mean vector. For each $v \neq 0$ suppose that (1.10) holds for all u satisfying $u'v = 0$. Then X has a spherical distribution.

The proof of Theorem 1 is given in Section 2. In Section 3, an alternative characterization of spherical distribution is given without assuming X has a mean vector.

The implications of Theorem 1 for linear models are the following. The fairly natural condition (1.7) concerning the error vector leads to a non-linear version of the Gauss-Markov Theorem as expressed by (1.6). However, (1.7) implies (1.10) which in turn implies that $\Sigma^{-\frac{1}{2}}\epsilon = X$ has a spherical distribution. Hence the error distribution must be elliptical (by definition, an elliptical distribution is a distribution obtained as a linear transformation of a spherical distribution). Therefore, the only error distributions for which (1.7) holds are the elliptical error distributions.

2. Main Theorem

Throughout this section, X is a random vector in R^n which has a mean vector. It is assumed that for each $v\neq 0$, $v\varepsilon R^n$, and for each $u\varepsilon R^n$ satisfying $u'v=0$, that

$$E(u'X|v'X) = 0 \qquad (2.1).$$

Let 0_n denote the group of nxn orthogonal matrices.

Theorem 1: When (2.1) holds, $L(X) = L(gX)$ for all $g\varepsilon 0_n$.

Proof: First recall that if U_1 and U_2 are real valued random variables with $E|U_1| < +\infty$, then the partial derivative

$$\frac{\partial}{\partial s_1} E \exp[i(s_1U_1+s_2U_2)] = iE\, U_1 \exp[i(s_1U_1+s_2U_2)] \qquad (2.2)$$

exists for all $s_1, s_2 \varepsilon R^1$. Furthermore, the condition $E(U_1|U_2) = 0$ implies that the partial derivative in (2.2) evaluated at $s_1=0$ is zero--that is

$$E\, U_1 \exp[i\, s_2U_2] = 0 \qquad (2.3)$$

for all $s_2 \varepsilon R^1$.

Let $\phi(t) = E \exp[i\, t'X]$ for $t\varepsilon R^n$. Since X has a mean vector, ϕ has a gradient given by

$$(\nabla\phi)(t) = i\, EX \exp[i\, t'X] \qquad (2.4)$$

Thus, for any $u,v\varepsilon R^n$, (2.4) yields

$$u'(\nabla\phi)(v) = i\, Eu'X \exp[i\, v'X].$$

For $u'v=0$, assumption (2.1) together with equation (2.3) ($U_1=u'X$, $U_2=v'X$, $s_2=1$) yields

$$u'\nabla\phi(v) = 0 \qquad (2.5)$$

for all $v\neq 0$ and u satisfying $u'v=0$.

To show that $L(X) = L(gX)$ for all $g\varepsilon 0_n$, it suffices to show that $\phi(t) = \phi(gt)$ for $t\varepsilon R^n$ and $g\varepsilon 0_n$. For $t=0$, this obviously holds, so fix

$t \neq 0$ and fix $g \varepsilon O_n$. Since t and gt have the same length, say $||t|| = r$, there is a continuously differentiable curve $c(\alpha)$, $\alpha \varepsilon (0,1)$, which satisfies $||c(\alpha)|| = r$ for all $\alpha \varepsilon (0,1)$, $c(\alpha_1) = t$ and $c(\alpha_2) = gt$ for some $\alpha_1, \alpha_2 \varepsilon (0,1)$. Because $||c(\alpha)||^2$ is constant, we have

$$(c(\alpha))' \dot{c}(\alpha) = 0 \quad , \quad \alpha \varepsilon (0,1), \tag{2.6}$$

so $c(\alpha)$ is perpendicular to the vector of derivatives $\dot{c}(\alpha)$. Using the chain rule, we have

$$\frac{d}{d\alpha} \phi(c(\alpha)) = [\dot{c}(\alpha)]' (\nabla\phi)(c(\alpha)) \tag{2.7}.$$

Since $c(\alpha) \neq 0$, (2.6) and (2.5) imply that (2.7) is zero for $\alpha \varepsilon (0,1)$ and hence that $\phi(c(\alpha))$ is constant for $\alpha \varepsilon (0,1)$. This implies that

$$\phi(t) = \phi(c(\alpha_1)) = \phi(c(\alpha_2)) = \phi(gt)$$

and the proof is complete.

3. An Alternative Characterization

In this section, we give an alternative characterization of the spherical distributions on R^n. Let X be a random vector in R^n with characteristic function ϕ. Fix a vector $u \epsilon R^n$, $||u|| = 1$ and let $g_u = I_n - 2uu'$ so g_u is an orthogonal matrix which maps u to -u and $g_u v = v$ for any v which is perpendicular to u. Let $P_u = uu'$ be the orthogonal projection onto span{u} and let $Q_u = I - P_u$. Our first task is to describe the implications of assuming that

$$L(P_u X | Q_u X) = L(-P_u X | Q_u X) \tag{3.1}$$

Notice that (3.1) is a stronger assumption than (1.9) when X has a mean. However, X is not assumed to have a mean in this section.

Lemma 3.1: When 3.1 holds,

$$L(X) = L(g_u X). \tag{3.2}$$

Proof: It suffices to verify that

$$\phi(t) = \phi(g_u t) \quad , \quad t \epsilon R^n. \tag{3.3}$$

Write $t = au+v$ for some $a \epsilon R^1$ and a unique v which satisfies $u'v = 0$. Then

$$g_u t = -au+v$$

so we need to show that

$$\phi(au+v) = \phi(-au+v). \tag{3.4}$$

Since $P_u u = u$, $P_u v = 0$, $Q_u u = 0$ and $Q_u v = v$, we have

$$\phi(-au+v) = E\exp[i\ (-au+v)'X] =$$

$$E\exp[i(-au+v)'(P_u X + Q_u X)] =$$

$$E\exp[-i\ au'P_u X + i\ v'Q_u X] =$$

$$E\{\exp[i\ v'Q_u X]\ E(\exp[i\ au'(-P_u X)] | Q_u X)\}.$$

Using (3.1), the above conditional expectation remains the same when $-P_uX$ is replaced by P_uX. Making this replacement and reading the above string of equalities in reverse order yields

$$\phi(-au+v) = \phi(au+v).$$

Hence (3.3) holds and the proof is complete.

Now, let

$$G = \{g | g\varepsilon 0_n, \ L(X) = L(gX)\} \tag{3.5}$$

If $L(X) = L(g_iX)$ for i=1,2, then $L(X) = L(g_1X) = L(g_1g_2X)$ so G is a subgroup of 0_n. Also, an easy continuity argument shows that G is closed. In fact, if S is any set of nxn orthogonal matrices such that $L(X) = L(gX)$ for all $g\varepsilon S$, then the group generated by S, say G(S), satisfies

$$L(X) = L(gX) \text{ for all } g\varepsilon G(S) \tag{3.6}$$

Proposition 3.1: If (3.1) holds for all $u\varepsilon R^n$, $||u|| = 1$, then X is spherical.

Proof: Let

$$S_o = \{g_u | u\varepsilon R^n, \ ||u|| = 1\}$$

with g_u as defined above. Also, let G_o be the group generated by S_o. When (3.1) holds for all u, $||u|| = 1$, then $S_o \subseteq G$ so $G_o \subseteq G$. However, every nxn orthogonal matrix is a product of a finite number of g_u's (This fact follows easily from the rotation angle representation of orthogonal matrices--see Theorem 1, p. 438 of Vilenkin (1968).). Thus, $0_n = G_o \subseteq G \subseteq 0_n$ so X is spherical. This completes the proof.

Under certain circumstances, it is possible to prove that X is spherical when (3.1) holds for some subset of u's, $||u|| = 1$. To discuss these circumstances, consider a set

$$B \subseteq \{u \ | \ ||u|| = 1\} \tag{3.7}$$

and assume that (3.1) holds for all $u \varepsilon B$. Let G_1 be the closed (in the topology of O_n) group generated by the set $\{g_u | u \varepsilon B\}$, so $G_1 \subseteq G$. If $G_1 = O_n$ then X is spherical, so we are interested in conditions so that $G_1 = O_n$. Before discussing such issues, it is necessary to first discuss reflection groups.

An nxn orthogonal matrix of the form $I - 2uu'$, $||u|| = 1$, is called a <u>reflection</u> because it reflects vectors across the hyperplane perpendicular to u. A group $H \subseteq O_n$ is called a <u>reflection group</u> if H is generated by some set of reflections. Recall that H is <u>reducible</u> if there exists a non-trivial subspace M of R^n such that $hM \subseteq M$ for all $h \varepsilon H$. If H is not reducible, H is called <u>irreducible</u>.

<u>Theorem</u>: (Eaton and Perlman (1977)). Let H be an irreducible reflection group. Then either H is a finite group or H is dense in O_n.

Since the reflection group G_1 defined above is closed, when G_1 is irreducible, either G_1 is finite or $G_1 = O_n$. A complete list of the finite irreducible reflection groups can be found in Theorem 5.3.1 of Benson and Grove (1971). When G_1 (irreducible) is not one of these, then $G_1 = O_n$ and X is spherical.

In order to apply the above argument to prove X is spherical, it must be verified that G_1 is irreducible and that G_1 is not finite (or not one of the groups given in Benson and Grove (1971)). Here is one useful condition which implies that G_1 is irreducible.

<u>Proposition 3.2</u>: Suppose B contains a set of vectors $\{x_1, x_2, \ldots, x_{n+1}\}$ and suppose every subset of size n of $\{x_1, \ldots, x_{n+1}\}$ is a linearly independent set in R^n. Then, G_1 is irreducible.

<u>Proof</u>: First observe that a non-trivial subspace M is invariant under a reflection g_u iff either $u \varepsilon M$ or $u \varepsilon M^{\perp}$. Now, assume that the non-trivial

subspace M is invariant under G_1. Then $M^{\perp}$ is also invariant under G_1 because $G_1 \subseteq O_n$. Thus, each x_i in $\{x_1, \ldots, x_{n+1}\}$ must be in M or in $M^{\perp}$. An easy dimension argument shows this is impossible because M is non-trivial. Hence G_1 is irreducible.

The verification that G_1 is not finite most often reduces to studing the angles between the vectors in B. For example, if $x_1, x_2 \varepsilon B$ and if the angle between x_1 and x_2 is an irrational multiple of π, then the group generated by g_{x_1} and g_{x_2} is infinite so G_1 is infinite.

References

Benson, C. and Grove, L. (1971). Finite Reflection Groups, Bogden and Quigley, Inc., Tarrytown-on-Hudson, New York.

Cambanis, S., Huang, S. and Simons, G. (1981). On the theory of elliptically contoured distributions. J. Mult. Anal. **11**, 368-385.

Eaton, M. L. and Perlman, M. (1977). Generating O_n with reflections. Pac. J. Math. **73**, No. 1, 73-80.

Eaton, M. L. (1983). The Gauss-Markov Theorem in multivariate analysis. To appear in Proceedings of Sixth International Symposium on Multivariate Analysis.

Kariya, T. and Toyooka, Y. (1983). Nonlinear versions of the Gauss-Markov Theorem and the GLSE. To appear in Proceedings of Sixth International Symposium on Multivariate Analysis.

Toyooka, Y. (1982). Prediction error in a linear model with estimated parameters. Biometrika, **69**, 453-459.

Vilenkin, N. (1968). Special Functions and the Theory of Group Representation. Published by American Math. Soc.